DOCUMENT RESUME

ED 347 168 TM 018 495

AUTHOR Bright, Elizabeth L.
TITLE Teachers' Views of Ethical Standardized Test Use.
PUB DATE Apr 92
NOTE 12p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 21-23, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; Elementary Education; *Elementary School Teachers; *Ethics; Factor Analysis; *Likert Scales; Scores; *Standardized Tests; Student Evaluation; Surveys; *Teacher Attitudes; *Test Construction; Test Use
IDENTIFIERS Curriculum Alignment; Teaching to the Test

ABSTRACT

In recent years, standardized test scores have assumed an increasingly important role in educational assessment. As the stakes involved in testing become higher, educators are forced to make decisions regarding the ethics involved in standardized test preparation and administration. The ethical continuum of test preparation practices as formulated by W. A. Mehrens and J. Kaminski (1989) is discussed. This continuum is instrumentalized in the form of scenarios pertaining to various standardized testing practices. A preliminary administration of a survey, developed to present these 27 scenarios was conducted with 102 elementary school teachers in the spring of 1991. Respondents had to rate the depicted behaviors on a four-point Likert scale ranging from very unethical to very ethical. The results indicate that the respondents do not view curriculum alignment activities on the same ethical behavior dimension as other practices that affect test scores. One table and 13 references are included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED347168

TEACHERS' VIEWS OF ETHICAL STANDARDIZED TEST USE

Elizabeth L. Bright
University of Florida

Paper presented at the annual meeting of the
National Council on Measurement in Education.
San Francisco, CA. April, 1992.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ELIZABETH L. BRIGHT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM018495

BEST COPY AVAILABLE

TEACHERS' VIEWS OF ETHICAL STANDARDIZED TEST USE

Elizabeth L. Bright
University of Florida

In recent years, standardized test scores have assumed an increasingly important role. As the stakes involved in testing become higher, educators are forced to make decisions regarding the ethics involved in practices related to standardized test preparation and administration. The ethical continuum of test preparation practices as formulated by Mehrens and Kaminski (1989) is discussed. This continuum is instrumentalized in the form of scenarios pertaining to various standardized testing practices. The results of a preliminary administration of the survey are discussed, as well as implications for future research in the area.

Introduction

Along with the dawn of the age of accountability, a new era of unethical test use and "test score pollution" has also arrived. Students as well as teachers and administrators are constantly undergoing performance evaluation. Whether it involves students competing for a scholarship, a teacher's salary increase to be based upon students' achievement, or the implementation or termination of a state or federally funded program, high stakes testing is affecting our nation's schools. (Shepard, 1990).

As standardized tests become increasingly high-stakes, these exams and their administration are increasingly scrutinized to determine the validity of the conclusions which we draw from them. Questions such as, "How closely did student preparation for the test parallel the content of the test?" are being asked as our standardized testing practices come under closer scrutiny.

Berk (1988) reported that, "...student achievement [as measured by test scores] is a criterion of teacher effectiveness in one-third of all statewide teacher incentive/ school incentive/career ladder programs." (p. 346). This emphasis on achievement has an effect on teaching practices. In a study of how standardized tests affect teachers, Smith (1991) found that "...teachers believe that scores are used against them... and do what is necessary to avoid low test scores." (p. 8).

There is a growing body of literature which describes how teachers go about avoiding low test scores. Haladyna, Nolen, and Haas (1991) discussed various ways that teachers may "pollute" test scores. These practices include: the teaching of test-taking skills, the promoting of student motivation for the test, the development of a curriculum designed to match the test, the presentation of items similar to those presented on the test, the erasure of stray marks, the dismissal of low-achieving students on tests days, and the assisting of students with responses. (p. 4).

Just how ethical are these behaviors and where is the line drawn between ethical and unethical test preparation? Mehrens and Kaminski (1989) devised a continuum of test preparation activities, where the following behaviors are arranged in a hierarchy from most ethical to most unethical:

1. General instruction on objectives not determined by looking at any set of published test objectives;

2. The teaching of test taking skills;

3. Instruction on objectives generated from looking at objectives measured by a variety of standardized tests;

4. Instruction based on objectives that specifically match those on the standardized or state test to be administered;

5. Instruction on specifically matched objectives where the practice (instruction) follows the same format as the test questions:

6. Practice (instruction) on a published parallel form of the same test; and

7. Practice (instruction) on the same test.

(From Mehrens and Kaminski, p. 16).

According to Mehrens and Kaminski, (1) above is always ethical while (6) and (7) are never ethical. Additionally, (2) "would typically be considered ethical." The "point where one crosses over from a legitimate to an illegitimate practice" would be somewhere between (3) and (5). (Mehrens and Kaminski, p. 16).

The premise on which these conclusion are based pertains to the inferences we make about test scores. As Mehrens (1988) has stated:

> For tests used to inform the public about educational quality or to inform parents and students about how the education of the student is progressing, the inference is almost always to a domain that is greater than the objectives tested. Thus, we are misleading the public when we limit instruction to the objectives tested.

Mehrens and Kaminski's continuum of ethical test preparation activities is a useful one. Such a continuum provides a framework for instrumentalizing behaviors in test preparation along an ethical/unethical continuum. Such an instrument could provide insight into how ethical teachers view certain test preparation activities to be. Furthermore, factors which might have an effect on teachers' ethical views might be analyzed and the results of this analysis could be used in future teacher preparation programs, testing programs, and perhaps even in test score interpretation.

Such an instrumentalization of the continuum is not found in the literature per se. However, Popham (1991) provided two evaluative standards to be applied by educators as to the appropriateness of test-preparation practices. Based on these standards, he investigated how educators viewed the appropriateness of five test-preparation practices. Popham concluded that generalized test-taking preparation and varied-format preparation were viewed as appropriate practices by educators. The three other preparation procedures studied were not as uniformly viewed. The present study may be viewed as an extension of Popham's work, in that it includes a large number of actual scenarios which might occur in testing which teachers are asked to rate on an ethical continuum.

Other studies have been done in the broader area of factors affecting teacher attitudes towards testing. For example, Green and Williams (1989) investigated the effect of grade level taught and training received in testing and measurement and how this affects teachers' attitudes towards testing and how useful they view testing to be. They concluded that grade level taught and the amount of training received had an effect on reported test use, and grade level taught was found to have an effect on attitudes towards testing.

While studies such as the one done by Green and Williams are useful in their own right, they do not directly investigate the more specific issue of teachers' views of ethics in standardized testing. In order to investigate this issue further, it was necessary to design an instrument pertaining specifically to ethics in testing. In order to demonstrate how such a study might be carried out, an instrument based largely on the work of Mehrens and Kaminski was formulated and administered to a group of teachers. Although these results are based on a small number (N=102), they do serve to illustrate the type of research which needs to be done in the area of ethics in standardized testing.

Purpose

The purpose of this study was to instrumentalize Mehrens and Kaminski's ethical test preparation continuum, thereby providing insight into how teachers view ethical behaviors in standardized testing situations. It was of interest to ascertain what the underlying factors of teachers' ethical perceptions might be. This was investigated through the use of factor analysis.

Instrument

The survey includes demographic items as well as 27 scenarios that might occur in standardized testing. Respondents were asked to rate the depicted behavior on a four-point Likert scale ranging from very unethical to very ethical. Typical items are of the following format:

> Mrs. Johnson first covers areas from the curriculum frameworks that she is sure are on the test, so that the students will be prepared in those areas.
>
> Mr. Bowes sets aside 45 minutes every Friday for test preparation activities such as "sustained silent reading" and team vocabulary games.
>
> While reviewing her third grade class's answer sheets, Mrs. Kelley erases stray pencil marks.

As Mr. Smith walks around the room during testing, he notices his best student has skipped a line on the answer sheet, which will in turn cause him/her to incorrectly record the remaining answer. He points this oversight out to him/her.

Items are rated on a Likert scale where a response of "1" indicates "very unethical" and "4" indicates "very ethical." (The complete survey is available by request).

Data Source

During the spring of 1991, the survey was administered to a convenience sample of 102 teachers from a large metropolitan northeastern school district of homogeneous socioeconomic status. Eleven percent of the teachers were male. Twelve percent of the teachers taught early childhood education, with the remaining 88 percent teaching elementary education. No secondary-level teachers were included, mainly because less annual testing occurs at the secondary level than at the elementary level. Although the sample size in this study was somewhat abbreviated, the main purpose here was to demonstrate how such an instrument could be used, and to provide a stepping stone to other more large-scale research.

Method

The 27 scenario items were factor analyzed using the responses obtained from the teachers. First, an exploratory factor analysis was performed on the data. Through the use of the Scree plot, Bartlett's X^2 test, and examination of the factor loading matrix, four factors were retained. (See Gorsuch (1983) for a thorough treatment of exploratory factor analysis). A Varimax rotation was then used to rotate the factors to obtain simple structure.

After obtaining the factor structures produced by the exploratory factor analysis, insight was gained into what could be the underlying factors for the scenarios, and a confirmatory factor analysis was performed using the LISREL computer program. It was hypothesized that each question loaded on one of four factors.

Results

Two models were fit to the data, namely: (1) the null model of 27 factors; and (2) the hypothesized four-factor model. The four hypothesized factors are: I) Practices which are clearly ethical or unethical; II) Practices which occur under extenuating circumstances; III) Practices which may be commonly encouraged; and IV) Practices regarding curriculur alignment/test preparation.

The factor loading matrix for the four-factor model was constrained so that each scenario loaded only on the factor that it was hypothesized to

represent. In Table 1, items representative of the hypothesized factor and their mean ratings are provided.

The factor loadings for each scenario are significant and large. That is, the majority of factor loadings are all at least twice as large as their standard error. The calculated X^2 for the null model, which hypothesizes 27 factors is 983.31 with 351 degrees of freedom. X^2 for the alternative four factor model is 477.31 with 318 degrees of freedom. The relative increase in goodness-of-fit of the four factor model over the null model is significant and indicates that the four factor model fits the data significantly better than the null model.

Unlike traditional hypothesis testing, when looking at the fit of a particular model, smaller values of X^2 are indications of good fit and are desirable, and although the calculated X^2 for the four factor model appears to be very large, Jöreskog and Sorbom (1981) suggest that X^2 values be used as a broad indication of fit, rather than as a formal test of a hypothesis. They suggest using a ratio of X^2 to degrees of freedom. Ratio values of greater than approximately two are indications of poor fit. Here, the ratio = 477.31/318 or 1.50 and indicates satisfactory fit of the four factor model.

It is important to note that the goodness of fit of this model must be interpreted with great caution. Here, LISREL was used primarily for purposes of discussion. Other methods may have been more appropriate. See for example, Muthén (1983, 1984, and 1987). Because the data are considered to be ordered categorical, Muthén would argue that the more appropriate method of analysis would involve the LISCOMP computer program. However, due to the preliminary nature of this study, the sample size was not nearly large enough to employ LISCOMP here.

Table 1
Hypothesized Factors, Representative Scenarios, and Mean Ratings.*

Factor I: Practices Which are Clearly Ethical or Unethical	
Mrs. Smith bases her class's test preparation on general objectives which have not been determined by looking at any set of published test objectives.	(Mean = 3.52)

* Mean rating is the average rating given by the teachers to a scenario, where a response of 1 indicates very unethical and 4 indicates very ethical.

Table 1 Continued

While reviewing the class's answer sheets, Mrs. Price notices her best student ended up on the wrong line because a line on the answer sheet was skipped. She adjusts the answers by moving them all up one line. (Mean = 1.81)

Mr. Jarrod gets a copy of a test the day before its administration, noting the trickier questions. The next morning, she helps prepare the students for the test by playing a game which includes a few of the trickier questions she knows are on the test. (Mean = 1.65)

Factor II: Practices Which Occur Under Extenuating Circumstances

At Mr. Mile's inner-city school, there is a great need for computers, books, and other materials. The principal has stressed to him confidentially that his class must perform well on the test, or the school will not get all of the funds it needs to buy materials. Following his principal's suggestion, he allows his class double the allotted time given by the test publisher to complete the test. (Mean = 1.26)

Some have called Mrs. Price a bad teacher. She is worried about being fired at the end of her second year because her students have notoriously performed poorly on tests. Fearing for her job, she allows her fifth grade class to use dictionaries and calculators on the exam. (Mean = 1.33)

At Mrs. Wolfe's school, her test scores are used to determine how much her raise will be next year. Because she needs the largest raise possible, she alters the answer sheets of the lowest scoring quarter of her class to increase their scores. (Mean = 1.14)

Table 1 Continued

Factor III: Practices Which May Be Commonly Encouraged

Mr. Price reads the set of test instructions to the class before administering the exam, stressing the words carefully and quickly.	(Mean = 3.56)
While reviewing her third grade class's answer sheets, Mrs. Legg erases stray pencil marks.	(Mean = 3.17)
As Mr. Smith walks around the room during testing, he notices his best student has skipped a line on the answer sheet, which will in turn cause him/her to incorrectly record the remaining answers. He points this oversight out to him/her.	(Mean = 3.25)

Factor IV: Practices Regarding Curriculur Alignment/Test Preparation.

Mrs. Wolfe bases her class's test preparation on a variety of standardized tests' objectives.	(Mean = 3.22)
As a first year teacher, Mr. Jones is advised by his principal that he should include test drills every day, and that all of the teachers in the school do this.	(Mean = 2.57)
Mrs. Johnson first covers areas from the curriculum frameworks that she is sure are on the test, so that the students will be prepared in those areas.	(Mean = 3.21)

Discussion and Conclusions

The purpose of this study was to provide insight into how ethical teachers viewed various behaviors related to standardized testing to be. Data obtained from the survey were factor analyzed and a four factor model was retained. The four factors underlying teachers' ethical perceptions of standardized testing scenarios are: I) Practices which are clearly ethical or unethical; II) Practices which occur under extenuating

circumstances; III) Practices which may be commonly encouraged; and IV) Practices regarding curriculur alignment/test preparation. The study conducted here can, at best, be regarded as a pilot study. Nevertheless, it provides a preliminary indication that teachers do not seem to view curricular alignment activities on the same ethical behavior dimension as other practices that affect test scores.

Because the study is based on a small sample without cross-validation and an instrument in need of further development, there is much more research to be done in this area. Future work in this area could include:

1. Expansion of the instrument to include the topics of excusing certain students on test days and providing rewards to students for good performance;

2. The inclusion of more demographic characteristics (such as number of years teaching, degree obtained, participation in a course on testing and measurement, socioeconomic level of the school, and ethnic composition of the student body) in order to ascertain whether or not these variables have any effect on teachers' ethical perceptions;

3. The administration of the survey to a much larger sample of teachers (a minimum of 300); and

4. Additional analyses, including hierarchical linear models to separate out effects of the school as a whole versus individual teacher characteristics on teachers' ethical beliefs regarding standardized testing.

Results of this future work will shed more insight into teachers' perceptions of ethical behaviors in standardized testing and what variables may contribute to those perceptions. These findings will have an impact for teacher training, including how inservice training and coursework taken by teachers may affect their ethical beliefs, as well as have an impact on the interpretations made regarding test scores obtained by schools that differ on given variables.

REFERENCES

Berk, R.A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. Journal of Personnel Evaluation in Education, 1, 345-363.

Gorsuch, R.L. (1983). Factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Green, K.E., & Williams, E.J. (1989, March). Standardized test use by classroom teachers: Effects of training and grade level taught. Paper presented at the Annual Meeting of the National council on Measurement in Education, San Francisco.

Haladyna, T.M., Nolen, S.B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. Educational Researcher, 20, 2-7.

Jöreskog, K.G., & Sorbom, D. (1981). Lisrel V - Analysis of linear structural relations by the method of maximum likelihood [User's Guide]. Chicago: International Educational Services.

Mehrens, W.A. (1988, June). Matching tests and curriculum: Is it really important? Paper presented at the ECS/CDE Assessment Conference, Boulder, CO.

Mehrens, W.A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? Educational Measurement: Issues and Practice, 8 (1), 14-22.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49 (1), 115-132.

Muthén, B. (1987). LISCOMP: Analysis of linear structural equations with a comprehensive measurement model [User's Guide]. Mooresville, IN: Scientific Software.

Popham, W.J. (1991). Appropriateness of teachers' test-preparation practices. Educational Measurement: Issues and Practice, 10 (4), 12-15.

Shepard, L.A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? Educational Measurement: Issues and Practices, 9, 15-22.

Smith, M.L. (1990). Put to the test: The effects of external testing on teachers. Educational Researcher, 20, 8-11.